Stockholms universitet
Statistiska institutionen
Raul Cano

SKRIVNINGSDATUM: 06-02-2018

Skriftlig tentamen i **Undersökningsmetodik** (4,5 hp), ingående som moment 1 i kursen **Regressionsanalys och undersökningsmetodik, 15 hp.**

Skrivtid: 5 timmar
Hjälpmedel: Miniräknare utan lagrade formler eller lagrad text. Vidhäftade formel- och tabellblad (obs! vidhäftas endast de tabellsidor som behövs för den här tentamen).
Tentamensgenomgång och återlämning: tisdagen den 20 februari, kl. 16.00 i B705.
Därefter kan skrivningarna hämtas på studentexpeditionen, plan 7 i B-huset.

Tentamen består av fem uppgifter som kan ge totalt 100 poäng. För betyget A gäller 90-100 p., för betyget B gäller 80-89 p., för betyget C gäller 70-79 p., för betyget D gäller 60-69 p., för betyget E gäller 50-59 p., för betyget Fx gäller 40-49 p. och för betyget F gäller 0-39 p. För detaljerade betygskriterier se kursbeskrivningen på kurshemsidan.
För full poäng på en uppgift krävs fullständiga och väl motiverade lösningar.

Uppgift 1: (20 poäng)
En population innehåller $N = 4$ element

$x_1 = 6, x_2 = 10, x_3 = 4, x_4 = 2$

Ur denna population skall man göra ett obundet slumpmässigt urval (OSU), om $n = 2$ element, utan återläggning.

a). Beräkna samplingfördelningen för den stokastiska variabeln s^2. (5 poäng)
b). Beräkna $E(s^2)$. (5 poäng)
c). Beräkna $V(s^2)$. (10 poäng)

Uppgift 2: (20 poäng)
Uppgiften handlar om den lilla republiken Slovenien. För att bestämma en skattning av befolkningen 31 dec 2017 gjordes ett obundet slumpmässigt urval utan återläggning av 5 kommuner bland alla Sloveniens 50 kommuner. Man tog reda på den definitiva befolkningen vid denna tidpunkt i dessa kommuner och man hade total tillgång på motsvarande uppgifter på befolkningen år 2002 (den lilla republiken hade genomfört en folkräkning år 2002). Total i den lilla republiken uppgick befolkningen 31 dec 2002 till 1 964 tusentals personer (alltså 1 964 000 personer). Följande resultat erhölls, alla siffror i tusentals personer:

Kommun	31 dec 2002	31 dec 2017
Gorizia E	20	18
Sava C	58	60
Karst N	30	30
Mura W	10	12
Savinja S	20	22

a). Beräkna regressionsskattningen av den totala befolkningen i Slovenien 31 dec 2017 med befolkning 31 dec 2002 som hjälpinformation. (5 poäng)

b). Beräkna den skattade variansen för regressionsskattningen av den totala befolkningen i Slovenien 31 dec 2017 med befolkning 31 dec 2002 som hjälpinformation. (10 poäng)

c). Beräkna ett 95%-igt konfidensintervall för den totala befolkningen i Slovenien 31 dec 2017 med hjälp av regressionsskattningen. (5 poäng)

Uppgift 3: (20 poäng)
I den lilla republiken Slovenien vill man göra en undersökning för att ta reda på andel studenter vid Ljubljana universitet som kommer att rösta på Kandidat XXX i det kommande presidentvalet. Kandidat XXX är en världsberömd och framgångsrik affärsman. Populationen har stratifierats i tre stratum efter akademitillhörighet. En stickprovsundersökning genomfördes för n=700 individer. Följande resultat erhölls då:

Stratum	N_i	n_i	p_i
Handelshögskolan	900	126	0,88
Fakultet för humaniora, utbildning och samhällsvetenskap	1000	140	0,47
Fakultet för Juridik, psykologi och socialt arbete	3100	434	0,06

Beräkna ett 95%-igt konfidensintervall för andel studenter vid Ljubljana universitet som kommer att rösta på Kandidat XXX i det kommande presidentvalet. (skattningen 5 poäng, skattade variansen 10 poäng, konfidensintervallet 5 poäng)

Uppgift 4: (20 poäng)
På en stor arbetsplats kan personalen göra läkarbesök på arbetstid och arbetsgivaren ersätter samtliga omkostnader som kan uppkomma. Arbetsgivaren önskar uppskatta den genomsnittliga kostnaden för denna typ av ersättning under det första kvartalet av innevarande år. Innevarande kvartal samlades in ett stickprov på n=100 anställda på ett slumpmässigt sätt (OSU utan återläggning) ur populationen av samtliga anställda (N=1000). Stickprovsmedelvärdet blev 17500 kr. och stickprovsvariansen blev 500.

Beräkna ett 95%-igt konfidensintervall för den genomsnittliga kostnaden i hela populationen. (skattningen 5 poäng, skattade variansen 10 poäng, konfidensintervallet 5 poäng)

Uppgift 5: (20 poäng)
a). Antag att vi har en population av 1000 hushåll och vill skatta P = andel hushåll i populationen som är enpersonshushåll. Hur stort urval måste vi dra om vi skall göra ett OSU utan återläggning bland alla enheter (hushåll) i populationen samt vill ha en felmarginal på högst 4 % och en konfidensgrad på 95 %? (10 poäng)

b). Antag att vi vill skatta P = andelen som skulle rösta med regeringsalliansen om det vore val idag. Hur stort urval måste vi dra om vi skall göra ett OSU utan återläggning bland "Svenska folket 16 år och äldre" samt vill ha en felmarginal på högst 1 % och en konfidensgrad på 95 %? (10 poäng)

Formelsamling undersökningsmetodik

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad \hat{\tau} = N\bar{X}$$

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}}{n-1} = \frac{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}{n(n-1)}$$

Beräkning av stickprovsstorlek:

$$n \geq \frac{N\sigma^2}{D^2(N-1) + \sigma^2}$$

Stratifierat urval:

$$\bar{X}_{st} = \sum_{i=1}^{L} W_i \bar{X}_i \qquad V(\bar{X}_{st}) = \sum_{i=1}^{L} W_i^2 V(\bar{X}_i) \qquad \text{där} \quad W_i = \frac{N_i}{N}$$

Optimal allokering:

$$n_i = n\frac{N_i\sigma_i}{\sum_{j=1}^{L} N_j\sigma_j}$$

Skattning av medelvärde samt proportion per element:

$$\bar{X}_{kvot} = \frac{\sum_{i=1}^{n} \tau_i}{\sum_{i=1}^{n} m_i} \qquad \bar{X}_{VVR} = N\frac{\bar{\tau}}{M} \qquad p_{kvot} = \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} m_i} \qquad P_{VVR} = N\frac{\bar{a}}{M}$$

$Punktskattning$ OSU	$Varians$ $m.å.$	$Variansskattning$ $m.å.$	$Varians$ $u.å.$	$Variansskattning$ $u.å.$
$\bar{X}$	$\frac{\sigma^2}{n}$	$\frac{s^2}{n}$	$\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$	$\frac{s^2}{n}\left(1-\frac{n}{N}\right)$
$\hat{\tau}$	$N^2 \cdot \frac{\sigma^2}{n}$	$N^2 \cdot \frac{s^2}{n}$	$N^2 \cdot \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$	$N^2 \cdot \frac{s^2}{n}\left(1-\frac{n}{N}\right)$
p	$\frac{P(1-P)}{n}$	$\frac{p(1-p)}{n-1}$	$\frac{P(1-P)}{n}\left(\frac{N-n}{N-1}\right)$	$\frac{p(1-p)}{n-1}\left(1-\frac{n}{N}\right)$
$\hat{A}$	$N^2 \cdot \frac{P(1-P)}{n}$	$N^2 \cdot \frac{p(1-p)}{n-1}$	$N^2 \cdot \frac{P(1-P)}{n}\left(\frac{N-n}{N-1}\right)$	$N^2 \cdot \frac{p(1-p)}{n-1}\left(1-\frac{n}{N}\right)$

Tillägg till formelsamling undersökningsmetodik

Skattning av τ_X. Urval OSU

$$\hat{\tau}_{krot} = \hat{R} \cdot \tau_Z = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} z_i} \cdot \tau_Z$$

$$\hat{V}\left(\hat{\tau}_{krot}\right) = N^2 \left(\frac{N-n}{nN}\right) \frac{\sum_{i=1}^{n} \left(x_i - \hat{R} z_i\right)^2}{n-1}$$

$$\sum_{i=1}^{n} \left(x_i - \hat{R} z_i\right)^2 = \sum_{i=1}^{n} x_i^2 + \hat{R}^2 \sum_{i=1}^{n} z_i^2 - 2\hat{R} \sum_{i=1}^{n} x_i z_i$$

Skattning av μ_X. Urval OSU.

$$\hat{\mu}_{reg} = \overline{x} + b\left(\mu_Z - \overline{z}\right)$$

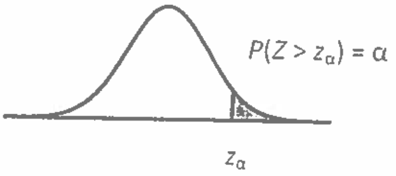
$$b = \frac{\sum_{i=1}^{n} \left(z_i - \overline{z}\right)\left(x_i - \overline{x}\right)}{\sum_{i=1}^{n} \left(z_i - \overline{z}\right)^2}$$

$$\hat{V}\left(\hat{\mu}_{reg}\right) = \left(\frac{N-n}{nN}\right)\left(\frac{1}{n-2}\right)\left[\sum_{i=1}^{n} \left(x_i - \overline{x}\right)^2 - b^2 \sum_{i=1}^{n} \left(z_i - \overline{z}\right)^2\right]$$

TABELL 2. Normalfördelningens kvantiler, standardiserad

$Z \in N(0, 1)$. Vilket värde har z_α om $P(Z > z_\alpha) = \alpha$ där α är en given sannolikhet.

Utnyttja även $\Phi(-z) = 1 - \Phi(z)$ för $P(Z \leq -z_\alpha)$.



α	z_α
0,25	0,6745
0,10	1,2816
0,05	1,6449
0,025	1,9600
0,010	2,3263
0,005	2,5758
0,0025	2,8070
0,0010	3,0902
0,0005	3,2905
0,00025	3,4808
0,00010	3,7190
0,00005	3,8906
0,000025	4,0556
0,000010	4,2649
0,000005	4,4172

© 2013 Michael Carlson, Statistiska inst., Sthlms univ

(5)

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 06/02/18

Sal: Brunnvikssalen

Tenta: Undersökningsmetodik

Kurs: Regressionsanalys och undersökningsmetodik

ANONYMKOD: OZR-NND

☒ Jag godkänner att min tenta får läggas ut anonymt på hemsidan som studentsvar.

OBS! SKRIV ÄVEN PÅ BAKSIDAN AV SKRIVBLADEN

Markera besvarade uppgifter med kryss

1	2	3	4	5	6	7	8	9	Antal inl. blad
X	X	X	X	X					6
Lär.ant. 20p	20p	20p	20p	20p					

9L

POÄNG	BETYG	Lärarens sign.
100p	A	RC

SU, STATISTIK

Skrivsal: Brunnsvikssalen   Anonymkod: OZR – NND   Blad nr: 1

1) 20p

Uppgift 1.

Populationen ifråga består av elementen
$x_1 = 6$
$x_2 = 10$
$x_3 = 4$
$x_4 = 2$

a) De möjliga urvalen (oordnade) är:

$x_1 \; x_2$	$\bar{x} = 8$	$s^2 = (-2)^2 + (2)^2 = 4+4 = 8$
$x_1 \; x_3$	$\bar{x} = 5$	$s^2 = \ldots = 1 + 1 = 2$
$x_1 \; x_4$	$\bar{x} = 4$	$s^2 = \ldots = 4 + 4 = 8$
$x_2 \; x_3$	$\bar{x} = 7$	$s^2 = \ldots = 9 + 9 = 18$
$x_2 \; x_4$	$\bar{x} = 6$	$s^2 = \ldots = 16 + 16 = 32$
$x_3 \; x_4$	$\bar{x} = 3$	$s^2 = \ldots = 1 + 1 = 2$

Den stokastiska variabeln s^2 (d.v.s. uppskattningen av populationsvariansen) definieras enligt: $s^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 / (n-1)$ där x_i här betecknar resp. individ i urvalet.

I vårt fall med $n=2$ fås $s^2 = \sum_{i=1}^{2}(x_i - \bar{x})^2$

Detta ger oss tredje kolumnen i tabellen ovan.

R

De möjliga urvalen i tabellen har samma sannolikhet $\frac{1}{6}$.

Två av urvalen ger samma s^2: 8, som därför får $P(s^2=8) = \frac{1}{3}$

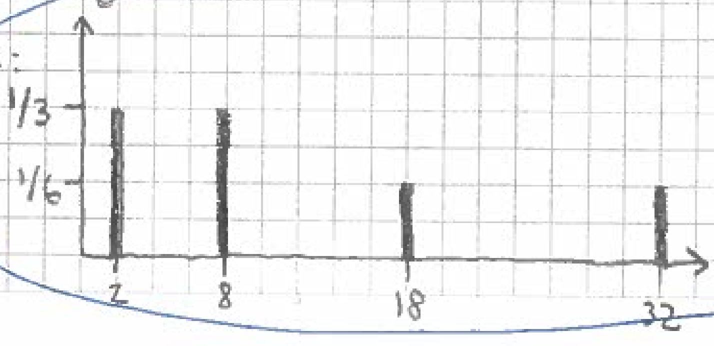
Två av urvalen ger samma s^2: 2, och därför fås $P(s^2=2) = \frac{1}{3}$

R

Samplingsfördelningen för s^2 blir således:

x:	P(x):
2:	$\frac{1}{3}$
8:	$\frac{1}{3}$
18:	$\frac{1}{6}$
32:	$\frac{1}{6}$

I grafisk form:



b) $E(s^2) = \sum_i s_i^2 \cdot P(i)$ där $i = 1, \ldots, 6$ löper över alla möjliga urval.

Således:

$E(s^2) = \frac{1}{6}(8+2+8+18+32+2) = 11{,}666\ldots$ R

R

Svar: $E(s^2) = 11\frac{2}{3} = 11{,}666\ldots \approx 11{,}67$ R

c) Variansen för s^2 definieras enligt:

$V(s^2) = \sum_i P(s_i^2) \cdot (s_i^2 - E(s^2))^2 = \{$Vi utnyttjar den framräknade samplingsfördelningen i deluppgift a).$\} =$

R

$= \frac{1}{3} \cdot \left(2 - 11\frac{2}{3}\right)^2 + \frac{1}{3}\left(8 - 11\frac{2}{3}\right)^2 + \frac{1}{6}\left(18 - 11\frac{2}{3}\right)^2 + \frac{1}{6}\left(32 - 11\frac{2}{3}\right)^2$

$\approx 31{,}14814811 + 4{,}481481465 + 6{,}685185199 + 68{,}90740741$

$\approx 111{,}22$ R

Svar: $V(s^2) \approx 111{,}22$ R

SU, STATISTIK

Skrivsal: Brunnsvikssalen Anonymkod: 0ZR-NND Blad nr: 2

Uppgift 2.

För att kunna utföra nödvändiga beräkningar kompletterar vi den givna tabellen ($\sum z_i = 138 \Rightarrow \bar{z} = 27,6$; $\sum x_i = 142 \Rightarrow \bar{x} = 28,4$):

Kommun	(31 dec 2002) z_i	(31 dec 2017) x_i	z_i^2	$z_i x_i$	x_i^2
Gorizia E	20	18	400	360	324
Sava C	58	60	3364	3480	3600
Karst N	30	30	900	900	900
Mura W	10	12	100	120	144
Savinja S	20	22	400	440	484
$\sum$	138	142	5164	5300	5452

a) För vår regressionsskattning behöver vi beräkna värdet på b enligt:

$$b = \frac{\sum_{i=1}^{n=5} (z_i - \bar{z})(x_i - \bar{x})}{\sum_{i=1}^{n=5} (z_i - \bar{z})^2} = \frac{\sum_{i=1}^{5} z_i x_i - 5\bar{x}\bar{z}}{\sum_{i=1}^{5} z_i^2 - 5\bar{z}^2} =$$

$$= \{\text{se tabell}\} = \frac{5300 - 5 \cdot 28,4 \cdot 27,6}{5164 - 5 \cdot 27,6^2} =$$

$$= \frac{1380,8}{1355,2} = 1,018890201$$ R

Regressionsskattningen $\hat{\mu}_{reg}$ av den genomsnittliga befolkningen R i en kommun den 31 december 2017 är $\boxed{\hat{\mu}_{reg} = \bar{x} + b(\mu_z - \bar{z})}$, där μ_z är populationsmedelvärdet för z. Men detta värde vet vi, eftersom folkmängden i Sloveniens 50 kommuner 31 dec 2002 var sammanlagt 1964 tusentals personer! Således: $\boxed{\mu_z = \frac{1964}{50} =}$ R

Uppgift 2a, forts.

Således: $\boxed{\mu_z = \frac{1964}{50} = 39,28}$ R

och vi får:

$$\hat{\mu}_{reg} = \bar{x} + b(\mu_z - \bar{z}) = 28,4 + 1,018890201(39,28 - 27,6) \approx$$

$$\boxed{\approx 40,3006376}$$ R

Den totala befolkningen skattas då till $\boxed{\hat{T}_x = \underline{50} \cdot \hat{\mu}_{reg} \approx}$ R

$\approx \boxed{2015,031877}$ R

R **Svar:** Regressionsskattningen av den totala befolkningen i Slovenien 31 dec 2017 är 2015 tusentals personer (2015000 pers.)

b) Den skattade variansen för den genomsnittliga kommunstorleken $\hat{\mu}_{reg}$ blir:

$$\underline{v(\hat{\mu}_{reg})} = \left(\frac{N-n}{nN}\right)\left(\frac{1}{n-2}\right)\left(\sum_{i=1}^{n}(x_i - \bar{x})^2 - b^2 \sum_{i=1}^{n}(z_i - \bar{z})^2\right) =$$

$$= \left\{\begin{matrix} n = 5 \\ N = 50 \end{matrix}\right\} = \frac{45}{250} \cdot \frac{1}{3}\left(\sum_{i=1}^{5} x_i^2 - 5\bar{x}^2 - b^2\left(\sum_{i=1}^{5} z_i^2 - 5\bar{z}^2\right)\right) \approx$$

$$\approx \left\{\text{Sätt in värden från tabellen}\right\} \approx 0,06\left(5452 - 5 \cdot 28,4^2 - 1,01889^2 \cdot\right.$$

$$\left. \cdot (5164 - 5 \cdot 27,6^2)\right) \approx \underline{0,738984603}$$ R

Den skattade variansen för den totala befolkningen blir:

$$\boxed{\hat{V}(T_x) = 50^2 \cdot v(\hat{\mu}_{reg}) \approx \underline{50^2 \cdot 0,738984603} \approx \underline{1847,46}}$$ R

Svar: Den skattade variansen för regressionsskattningen av den totala befolkningen 31 dec 2017 är $\underline{\hat{V}(T_x) \approx 1847,46}$ R där T_x anges i tusentals personer.

Obs! För deluppgift c: se blad 3.

SU, DEPARTMENT OF STATISTICS

Room: Brunnsvikssalen Anonymous code: OZR-NND Sheet number: 3

Uppgift 2c

c) För ett 95% konfidensintervall använder vi $z_{0,025} = 1,9600$ och får intervallgränserna $\hat{T}_x \pm z_{0,025} \cdot \sqrt{\hat{V}(\hat{T}_x)}$

eller 2015,03 ± 1,96 · 42,9821

Annorlunda uttryckt blir intervallet (1931; 2099) R

R

Svar: Ett 95% konfidensintervall för den totala befolkningen i Slovenien 31 dec 2017 är (1931; 2099) där befolkningen anges i tusentals personer.

R

SU, STATISTIK

Skrivsal: Brunnsvikssalen Anonymkod: OZR-NND Blad nr: 4

3) 20p

Uppgift 3

Vi uppskattar $V(p_i)$ med uttrycket $\frac{p_i(1-p_i)}{n_i-1}\left(1-\frac{n_i}{N_i}\right)$, där i betecknar stratum. Med data från tabellen fås:

$V(p_1) = \frac{0,88\cdot(1-0,88)}{126-1}\left(1-\frac{126}{900}\right) \approx 0,000726528$ R

$V(p_2) = \frac{0,47\cdot(1-0,47)}{140-1}\left(1-\frac{140}{1000}\right) \approx 0,001541194$ R

$V(p_3) = \frac{0,06\cdot(1-0,06)}{434-1}\left(1-\frac{434}{3100}\right) \approx 0,000112018$ R

- $V(p_{st}) = \sum_{i=1}^{3}\left(\frac{N_i}{N}\right)^2 V(p_i) = \{N = 5000 = N_1+N_2+N_3\} =$

$= \left(\frac{900}{5000}\right)^2\cdot 0,000726528 + \left(\frac{1000}{5000}\right)^2\cdot 0,001541194 + \left(\frac{3100}{5000}\right)^2\cdot 0,000112018 \approx$

$\approx 0,00002354 + 0,00006 1648 + 0,00004306 \approx$

$\approx 0,000128248$ R

$\Rightarrow \sqrt{V(p_{st})} \approx 0,011324651$ R

- Vår punktskattning av p_{st} blir vidare $p_{st} = \sum_{i=1}^{3}\frac{N_i}{N}p_i =$

$= \frac{900}{5000}\cdot 0,88 + \frac{1000}{5000}\cdot 0,47 + \frac{3100}{5000}\cdot 0,06 = 0,2896$ R

För ett 95% konfidensintervall använder vi $Z_{0,025} = 1,9600$ och får intervallgränserna $p_{st} \pm Z_{0,025}\cdot\sqrt{V(p_{st})}$ eller $0,2896 \pm 1,96\cdot 0,0113247$

Annorlunda uttryckt blir intervallet $(0,267; 0,312)$ R

R

SU, STATISTIK

Skrivsal: Brunnsvikssalen Anonymkod: OZR-NND Blad nr: 5

4) 2Op

Uppgift 4.

Givna data:

$n = 100$

$N = 1000$

$\bar{x} = 17500$ (kr)

$s^2 = 500$

Låt x_i beteckna kostnaden i kr för individ i.

Vår skattning av den genomsnittliga kostnaden i population blir stickprovsmedelvärdet $\bar{x} = 17500$ R

Vår variansskattning blir $V(\bar{x}) = \frac{s^2}{n}\left(1 - \frac{n}{N}\right) = \frac{500}{100} \cdot \left(1 - \frac{100}{1000}\right) =$ R

$= 4,5$ R

För ett 95% konfidensintervall använder vi $z_{0,025} = 1,9600$ och får intervallgränserna $\bar{x} \pm z_{0,025} \cdot \sqrt{V(\bar{x})}$ eller $17500 \pm 1,96 \cdot 2,12132$ R

Annorlunda uttryckt blir intervallet $(17495,84; 17504,16)$.

R

SU, STATISTIK

Skrivsal: Brunnsvikssalen Anonymkod: OZR-NND Blad nr: 6

5) 20p

Uppgift 5.

a) Kravet att felmarginalen ej får överstiga 4 % vid konfidensgraden 95 % kan uttryckas som:

$$*)\ 1{,}9600 \cdot \sqrt{V(p)} \le 0{,}04$$

Vi vill gardera oss mot det "värsta" scenariot att $p = 0{,}5$. Anta därför $p = 0{,}5$. Då fås (med urvalsstorleken n):

$$V(p) = \frac{0{,}5\cdot(1-0{,}5)}{n}\left(\frac{N-n}{N-1}\right) = \{N = 1000\} =$$

$$= \frac{0{,}25}{n}\left(\frac{1000-n}{999}\right)$$

Insättning i $*)$ ovan ger olikheten:

R

$$1{,}9600\cdot\sqrt{\frac{0{,}25}{n}\left(\frac{1000-n}{999}\right)} \le 0{,}04 \Rightarrow$$

$$\Rightarrow \sqrt{\frac{999n}{0{,}25(1000-n)}} \ge \frac{1{,}96}{0{,}04} \Rightarrow \frac{999n}{0{,}25(1000-n)} \ge 2401 \Rightarrow$$

$$\Rightarrow 999n \ge 600250 - 600{,}25n \Rightarrow 1599{,}25n \ge 600250$$

$$\Rightarrow \boxed{n \ge 375{,}33219}$$ R

Svar: Vi måste dra ett urval av 376 hushåll. R

b) Kravet att felmarginalen ej får överstiga 1 % vid konfidensgraden 95 % kan uttryckas som:

*) $1{,}9600 \cdot \sqrt{V(p)} \leq 0{,}01$

För att hantera det "värsta" scenariot att $P = 0{,}5$, antar vi just $P = 0{,}5$

Variansen för skattningen p blir då, med urvalsstorleken n,

$$V(p) = \frac{P(1-P)}{n}\left(\frac{N-n}{N-1}\right) = \frac{0{,}5 \cdot 0{,}5}{n}\left(\frac{N-n}{N-1}\right) =$$

$= \{$ Anta att N är "stort". Det är ett rimligt antagande för populationen "Svenska folket 16 år och äldre". Då kan vi approximera $\frac{N-n}{N-1} \approx \frac{N-n}{N} \approx \{\text{Anta } N \gg n\} \approx 1 \} =$

$$= \frac{0{,}25}{n}$$

Insättning i *) ger nu olikheten:

$$1{,}9600 \cdot \sqrt{\frac{0{,}25}{n}} \leq 0{,}01 \Rightarrow \sqrt{\frac{n}{0{,}25}} \geq \frac{1{,}9600}{0{,}01} \Rightarrow$$

$$\Rightarrow \frac{n}{0{,}25} \geq 38416 \Rightarrow n \geq 9604$$ R

Svar: Vi måste dra ett urval av 9604 personer. R